



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A support and density theorem for Markovian rough paths

Citation for published version:

Chevyrev, I & Ogorodnik, M 2018, 'A support and density theorem for Markovian rough paths', *Electronic journal of probability*, vol. 23, 56. <https://doi.org/10.1214/18-EJP184>

Digital Object Identifier (DOI):

[10.1214/18-EJP184](https://doi.org/10.1214/18-EJP184)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Electronic journal of probability

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A SUPPORT AND DENSITY THEOREM FOR MARKOVIAN ROUGH PATHS

ILYA CHEVYREV AND MARCEL OGRODNIK

ABSTRACT. We establish two results concerning a class of geometric rough paths \mathbf{X} which arise as Markov processes associated to uniformly subelliptic Dirichlet forms. The first is a support theorem for \mathbf{X} in α -Hölder rough path topology for all $\alpha \in (0, 1/2)$, which proves a conjecture of Friz–Victoir [FV10]. The second is a Hörmander-type theorem for the existence of a density of a rough differential equation driven by \mathbf{X} , the proof of which is based on analysis of (non-symmetric) Dirichlet forms on manifolds.

1. INTRODUCTION

Consider a symmetric Dirichlet form on $L^2(\mathbb{R}^d, \lambda)$

$$(1.1) \quad \mathcal{E}(f, g) = \int_{\mathbb{R}^d} \sum_{i,j=1}^d a^{i,j}(\partial_i f)(\partial_j g) d\lambda,$$

where λ is the Lebesgue measure and a is a measurable, uniformly elliptic function taking values in the space of symmetric $d \times d$ matrices (we make our set-up precise in Section 1.1). It is well-known that there exists a symmetric Markov process \mathbf{X} in \mathbb{R}^d associated with \mathcal{E} ; see [FOT11] for a general construction of \mathbf{X} and [Str88] for fundamental analytic properties of \mathcal{E} .

We are interested in differential equations of the form

$$(1.2) \quad d\mathbf{Y}_t = V(\mathbf{Y}_t)d\mathbf{X}_t, \quad \mathbf{Y}_0 = y_0 \in \mathbb{R}^e,$$

driven by \mathbf{X} along vector fields $V = (V_1, \dots, V_d)$ on \mathbb{R}^e . When a is taken sufficiently smooth, the process \mathbf{X} can be realised as a semi-martingale for which the classical framework of Itô gives meaning to the equation (1.2). However for irregular functions a , this is no longer the case, and (1.2) falls outside the scope of Itô calculus.

One of the applications of Lyons' theory of rough paths [Lyo98] has been to give meaning to differential equations driven by processes outside the range of semi-martingales. One viewpoint of rough paths theory is that it factors the problem of solving equations of the type (1.2) into first enhancing \mathbf{X} to a rough path by appropriately defining its iterated integrals (which is typically done through stochastic means), and then solve (1.2) deterministically.

Probabilistic methods to enhance the Markov process \mathbf{X} to a rough path and the study of its fundamental properties appear in [LS99, BHL02, Lej06, Lej08], where primarily the forward-backward martingale decomposition is used to show

2010 *Mathematics Subject Classification.* Primary 60H10; Secondary 60G17.

Key words and phrases. Markovian rough paths, support in Hölder topology, Hörmander's theorem.

existence of the stochastic area. A somewhat different approach, which we follow here, is taken in [FV08] where the authors define \mathbf{X} directly as a diffusion on the free nilpotent Lie group $G^N(\mathbb{R}^d)$ (in particular the iterated integrals are given directly in the construction). One can show that in the situation mentioned at the start, the two methods give rise to equivalent definitions of rough paths. The latter construction in fact yields further flexibility in that the evolution of \mathbf{X} can depend in a non-trivial way on its higher levels (its iterated integrals). Note that this is a common feature with Lévy rough paths studied in [FS17, Che18]. Markovian rough paths have also recently been investigated in [CO17, CL16] in connection with the accumulated local p -variation functional and the moment problem for expected signatures.

The goal of this paper is to contribute two results to the study of Markovian rough paths in the sense of [FV08]. Our first contribution (Theorem 2.11) answers in the positive a conjecture about the support of \mathbf{X} in α -Hölder rough path topology. Such a support theorem appeared in [FV08] for $\alpha \in (0, 1/6)$, and was improved to $\alpha \in (0, 1/4)$ in [FV10] where it was conjectured to hold for $\alpha \in (0, 1/2)$ in analogy to enhanced Brownian motion. Comparing our situation to the case of Gaussian rough paths, where such support theorems are known with sharp Hölder exponents (see e.g., [FV10, Sec. 15.8], and [FGGR16] for recent improvements), the difficulty of course lies in the lack of a Gaussian structure, in particular the absence of a Cameron-Martin space.

Our solution to this problem relies almost entirely on elementary techniques. Indeed, we first show that any stochastic process (taking values in a Polish space) admits explicit lower bounds on the probability of keeping a small α -Hölder norm, provided that it satisfies lower and upper bounds on certain transition probabilities comparable to Brownian motion. This is made precise by conditions (1) and (2) and Theorem 2.5. We then verify these conditions for the translated rough path $T_h(\mathbf{X})$ (which is in general non-Markov, see Remark 2.8) for any $h \in W^{1,2}([0, T], \mathbb{R}^d)$ using heat kernel estimates of \mathbf{X} (we also note that, just like for enhanced Brownian motion, all relevant constants depend on h only through $\|h\|_{W^{1,2}}$).

As usual, in combination with the continuity of the Itô-Lyons map from rough paths theory, an immediate consequence of improving the Hölder exponent in the support theorem for \mathbf{X} is a stronger Stroock-Varadhan support theorem (in α -Hölder topology) for the solution \mathbf{Y} to the rough differential equation (RDE) (1.2) along with the lower regularity assumptions on the driving vector fields V (Lip^2 instead of Lip^4).

Our second contribution (Theorem 3.4 and its Corollary 3.8) may be seen as a non-Gaussian Hörmander-type theorem, and provides sufficient conditions on the driving vector fields $V = (V_1, \dots, V_d)$ under which the solution to the RDE (1.2) admits a density with respect to the Lebesgue measure on \mathbb{R}^e . Once again, while this result is reminiscent of density theorems for RDEs driven by Gaussian rough paths (e.g., [BH07, CF10, CHLT15]), the primary difference in our setting is that methods from Malliavin calculus are no longer available due to the lack of a Gaussian structure.

We replace the use of Malliavin calculus by direct analysis of (non-symmetric) Dirichlet forms on manifolds. Indeed, we identify conditions under which the couple (\mathbf{X}, \mathbf{Y}) admits a density on its natural state-space, and conclude by projecting to \mathbf{Y} . We note however that our current result gives no quantitative information

about the density beyond its existence (not even for the couple (\mathbf{X}, \mathbf{Y})), and we strongly suspect that the method can be improved to yield further information (particularly L^p bounds and regularity results in the spirit of the De Giorgi–Nash–Moser theorem).

1.1. Notation. Throughout the paper, we adopt the convention that the domain of a path $\mathbf{x} : [0, T] \rightarrow E$, for $T > 0$ and a set E , is extended to all of $[0, \infty)$ by setting $\mathbf{x}_t = \mathbf{x}_T$ for all $t > T$. For a metric space (E, d) , $r \geq 0$, and $x \in E$, we denote the ball $B(x, r) = \{y \in E \mid d(x, y) \leq r\}$.

We let $G = G^N(\mathbb{R}^d)$ denote the step- N free nilpotent Lie group over \mathbb{R}^d for some $N \geq 2$, and let U_1, \dots, U_d be a set of generators for its Lie algebra $\mathfrak{g} = \mathfrak{g}^N(\mathbb{R}^d)$, which we identify with the space of left-invariant vector fields on G . We equip \mathbb{R}^d with the inner product for which U_1, \dots, U_d form an orthonormal basis upon canonically identifying \mathbb{R}^d with a subspace of \mathfrak{g} .

We equip G with the corresponding Carnot–Carathéodory metric d . Let 1_G denote the identity element of G and let λ denote the Haar measure on G normalised so that $\lambda(B(1_G, 1)) = 1$.

For $\Lambda > 0$, let $\Xi(\Lambda) = \Xi^{N,d}(\Lambda)$ denote the set of measurable functions a on G which take values in the space of symmetric $d \times d$ matrices and which are sub-elliptic in the following sense:

$$\Lambda^{-1}|\xi|^2 \leq \langle \xi, a(x)\xi \rangle \leq \Lambda|\xi|^2, \quad \forall \xi \in \mathbb{R}^d, \quad \forall x \in G.$$

For $a \in \Xi(\Lambda)$, we define the associated Dirichlet form $\mathcal{E} = \mathcal{E}^a$ on $L^2(G, \lambda)$ for all $f, g \in C_c^\infty(G)$ by

$$(1.3) \quad \mathcal{E}(f, g) = \int_G \sum_{i,j} a^{i,j}(U_i f)(U_j g) d\lambda.$$

We let $\mathbf{X} = \mathbf{X}^{a,x}$ denote the Markov diffusion on G associated to \mathcal{E} with starting point $\mathbf{X}_0 = x \in G$. We recall that the sample paths of \mathbf{X} are a.s. geometric α -Hölder rough paths for all $\alpha \in (0, 1/2)$, and when $a(x)$ depends only on the level-1 projection $\pi_1(x) \in \mathbb{R}^d$ of $x \in G$, \mathbf{X} serves as the natural rough path lift of the Markov diffusion associated to the Dirichlet form (1.1) on $L^2(\mathbb{R}^d)$ discussed earlier. For further details, we refer to [FV10].

Remark 1.1. Throughout the paper we assume the symmetric Dirichlet form (1.3) is defined on the Hilbert space $L^2(G, \lambda)$ so that \mathbf{X} is symmetric with respect to λ . As pointed out in [CO17], it is natural to also consider \mathcal{E} defined over $L^2(G, \mu)$ for a measure $\mu(dx) = v(x)\lambda(dx)$, $v \geq 0$. While for simplicity we only work with \mathcal{E} defined on $L^2(G, \lambda)$, we note that appropriate assumptions of v and a Girsanov transform (see, e.g., [Fit97]) can be used to relate the results of this paper to this more general setting.

2. SUPPORT THEOREM

2.1. Restricted Hölder norms. We first record some deterministic results on Hölder norms which will be used in the sequel. Throughout this section, let (E, d) be a metric space, $\alpha \in (0, 1]$, $T > 0$, and $\mathbf{x} \in C([0, T], E)$ a continuous path. Let \square denote any of the relations $<, \leq, =, \geq, >$, and consider the quantity

$$\|\mathbf{x}\|_{\alpha\text{-Höl}, \square_\varepsilon; [s, t]} = \sup_{u, v \in [s, t], |u-v| \square_\varepsilon} \frac{d(\mathbf{x}_u, \mathbf{x}_v)}{|u-v|^\alpha},$$

where we set $\|\mathbf{x}\|_{\alpha\text{-H\"ol}, \square_\varepsilon; [s, t]} = 0$ if the set $\{(u, v) \in [s, t]^2 \mid |u - v| \square_\varepsilon\}$ is empty.

Definition 2.1. For $\varepsilon, \gamma > 0$ and $s \in [0, T]$, define the times $(\tau_n^{\varepsilon, \gamma, s})_{n \geq 0} = (\tau_n)_{n \geq 0}$ by $\tau_0 = s$ and for $n \geq 1$

$$\tau_n = \inf\{t > \tau_{n-1} \mid \|\mathbf{x}\|_{\alpha\text{-H\"ol}, \geq \varepsilon; [\tau_{n-1}, t]} \geq \gamma\}.$$

We call any such τ_n a *H\"older stopping time* of \mathbf{x} .

Lemma 2.2. Let $\varepsilon, \gamma > 0$ and $s = 0$, and suppose that for some $c > 0$

$$(2.1) \quad \sup_{t \in [\tau_n, \tau_n + \varepsilon]} d(\mathbf{x}_{\tau_n}, \mathbf{x}_t) < c, \quad \forall n \geq 0.$$

Then $\|\mathbf{x}\|_{\alpha\text{-H\"ol}, = \varepsilon; [0, T]} < \tilde{\gamma} := (3c\varepsilon^{-\alpha}) \vee (4\gamma + c\varepsilon^{-\alpha})$.

Proof. For $n \geq 1$ and $t \in [\tau_n - \varepsilon, \tau_n]$, we have one of the following three mutually exclusive cases: (a) $\tau_n = \tau_{n-1} + \varepsilon$, (b) $\tau_n \in (\tau_{n-1} + \varepsilon, \tau_{n-1} + 2\varepsilon]$ and $t \in [\tau_{n-1}, \tau_{n-1} + \varepsilon]$, or (c) $t > \tau_{n-1} + \varepsilon$. In case (a), (2.1) implies that $d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) < 2c$. In case (b), $d(\mathbf{x}_{\tau_n}, \mathbf{x}_{\tau_{n-1}}) \leq \gamma(2\varepsilon)^\alpha$ and (2.1) implies that $d(\mathbf{x}_t, \mathbf{x}_{\tau_{n-1}}) < c$, so that

$$d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) < c + \gamma(2\varepsilon)^\alpha \leq (2c) \vee (4\gamma\varepsilon^\alpha).$$

In case (c), we have $d(\mathbf{x}_{t-\varepsilon}, \mathbf{x}_t) \leq \gamma\varepsilon^\alpha$ and $d(\mathbf{x}_{t-\varepsilon}, \mathbf{x}_{\tau_n}) \leq \gamma(2\varepsilon)^\alpha$, so that

$$d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) \leq \gamma\varepsilon^\alpha + \gamma(2\varepsilon)^\alpha \leq 3\gamma\varepsilon^\alpha.$$

Hence, in all three cases

$$(2.2) \quad d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) < (2c) \vee (4\gamma\varepsilon^\alpha).$$

Consider now

$$\tau = \inf\{t > 0 \mid \|\mathbf{x}\|_{\alpha\text{-H\"ol}, = \varepsilon; [0, T]} = \tilde{\gamma}\}.$$

Note that $\|\mathbf{x}\|_{\alpha\text{-H\"ol}, = \varepsilon; [0, T]} \geq \tilde{\gamma} \Leftrightarrow \tau < \infty$. Arguing by contradiction, suppose that $\tau < \infty$, which means that $d(\mathbf{x}_{\tau-\varepsilon}, \mathbf{x}_\tau) = \tilde{\gamma}\varepsilon^\alpha$. Consider the largest n for which $\tau_n \leq \tau$. Observe that $\tau_n \in [\tau - \varepsilon, \tau]$, since otherwise $d(\mathbf{x}_{\tau-\varepsilon}, \mathbf{x}_\tau) < \gamma\varepsilon^\alpha$, which is a contradiction since $\tilde{\gamma} > \gamma$. It follows from (2.1) that $d(\mathbf{x}_{\tau_n}, \mathbf{x}_\tau) \leq c$, and therefore by (2.2) and the triangle inequality

$$d(\mathbf{x}_{\tau-\varepsilon}, \mathbf{x}_\tau) < c + (2c) \vee (4\gamma\varepsilon^\alpha) = \tilde{\gamma}\varepsilon^\alpha,$$

which is again a contradiction. \square

Lemma 2.3. Suppose that $\|\mathbf{x}\|_{\alpha\text{-H\"ol}, = 2^{-n}\varepsilon; [0, T]} \leq \gamma$ for every $n > N \in \mathbb{Z}$. Then

$$\|\mathbf{x}\|_{\alpha\text{-H\"ol}, < 2^{-N}\varepsilon; [0, T]} \leq \frac{\gamma}{1 - 2^{-\alpha}}.$$

Proof. Consider $(t-s)/\varepsilon \in (0, 2^{-N})$ with binary representation $(t-s)/\varepsilon = \sum_{n=m}^{\infty} c_n 2^{-n}$ with $c_n \in \{0, 1\}$, $m > N$, and $c_m = 1$. It follows that

$$d(\mathbf{x}_s, \mathbf{x}_t) \leq \gamma \sum_{n=m}^{\infty} \varepsilon^\alpha c_n 2^{-n\alpha}.$$

Since $2^{-m} \leq (t-s)/\varepsilon$, we have $\varepsilon^\alpha 2^{-n\alpha} \leq 2^{\alpha(m-n)}(t-s)^\alpha$. Hence

$$d(\mathbf{x}_s, \mathbf{x}_t) \leq \gamma \sum_{n=m}^{\infty} 2^{\alpha(m-n)}(t-s)^\alpha = \frac{\gamma(t-s)^\alpha}{1 - 2^{-\alpha}}.$$

\square

Lemma 2.4. *Suppose there exist $x \in E$ and $r > 0$ such that for all integers $k \geq 0$, $\mathbf{x}_{k\varepsilon} \in B(x, r)$ and $\|\mathbf{x}\|_{\alpha\text{-H\"ol}; \leq \varepsilon; [k\varepsilon, (k+1)\varepsilon]} \leq \gamma$. Then*

$$\|\mathbf{x}\|_{\alpha\text{-H\"ol}; [0, T]} \leq 2\gamma + 2r\varepsilon^{-\alpha}.$$

Proof. Consider $0 \leq s < t \leq [0, T]$, and denote $s \in [k\varepsilon, (k+1)\varepsilon]$, $t \in [n\varepsilon, (n+1)\varepsilon]$. If $k = n$ there is nothing to prove, so suppose $k < n$. If $|t - s| \leq \varepsilon$, so that $n = k+1$, then

$$d(\mathbf{x}_s, \mathbf{x}_t) \leq d(\mathbf{x}_s, \mathbf{x}_{n\varepsilon}) + d(\mathbf{x}_{n\varepsilon}, \mathbf{x}_t) \leq \gamma 2^{1-\alpha} |t - s|^\alpha.$$

Finally, if $|t - s| > \varepsilon$ then since $\mathbf{x}_{k\varepsilon}, \mathbf{x}_{n\varepsilon} \in B(x, r)$, it follows that

$$\begin{aligned} |t - s|^{-\alpha} d(\mathbf{x}_s, \mathbf{x}_t) &\leq |t - s|^{-\alpha} (d(\mathbf{x}_{k\varepsilon}, \mathbf{x}_s) + d(\mathbf{x}_{k\varepsilon}, \mathbf{x}_{n\varepsilon}) + d(\mathbf{x}_{n\varepsilon}, \mathbf{x}_t)) \\ &\leq |t - s|^{-\alpha} (2\varepsilon^\alpha \gamma + 2r) \\ &\leq 2\gamma + 2r\varepsilon^{-\alpha}. \end{aligned}$$

□

2.2. Positive probability of small Hölder norm. Suppose now (E, d) is a Polish space. In this section, we give conditions under which an E -valued process has an explicit positive probability of keeping a small Hölder norm. We fix $\alpha \in (0, 1/2)$, a terminal time $T > 0$, and an E -valued stochastic process \mathbf{X} adapted to a filtration $(\mathcal{F}_t)_{t \in [0, T]}$.

Consider the following conditions:

- (1) There exists $C_1 > 0$ such that for every $c, \varepsilon > 0$, and every Hölder stopping time τ of \mathbf{X} , a.s.

$$\mathbb{P} \left[\sup_{t \in [\tau, \tau + \varepsilon]} d(\mathbf{X}_\tau, \mathbf{X}_t) > c \mid \mathcal{F}_\tau \right] \leq C_1 \exp \left(\frac{-c^2}{C_1 \varepsilon} \right).$$

- (2) There exist $c_2, C_2 > 0$ and $x \in E$ such that for every $s \in [0, T]$ and $\varepsilon \in (0, T - s]$, a.s.

$$\mathbb{P}[\mathbf{X}_{s+\varepsilon} \in B(x, C_2 \varepsilon^{1/2}) \mid \mathcal{F}_s] \geq c_2 \mathbf{1}\{\mathbf{X}_s \in B(x, C_2 \varepsilon^{1/2})\}.$$

Roughly speaking, the first condition states that the probability of large fluctuations of \mathbf{X} over small time intervals should have the same Gaussian tails as that of a Brownian motion, while the second condition bounds from below the probability that $\mathbf{X}_{s+\varepsilon}$ is in a ball of radius $\sim \varepsilon^{1/2}$ given that \mathbf{X}_s was in the same ball.

Theorem 2.5. *Assume conditions (1) and (2). Fix x as in (2). Then there exist $C_{2.5}, c_{2.5} > 0$, depending only on C_1, c_2, C_2, α, T , such that for every $\gamma > 0$, a.s.*

$$\mathbb{P} [\|\mathbf{X}\|_{\alpha\text{-H\"ol}; [0, T]} < \gamma \mid \mathcal{F}_0] \geq C_{2.5}^{-1} \exp \left(\frac{-C_{2.5}}{\gamma^{2/(1-2\alpha)}} \right) \mathbf{1}\{\mathbf{X}_0 \in B(x, c_{2.5} \gamma^{1/(1-2\alpha)})\}.$$

Lemma 2.6. *Assume condition (1). Then there exists $C_{2.6} > 0$, depending only on C_1 and α , such that for all $0 \leq s < t \leq T$ and $\varepsilon \in (0, t - s]$, a.s.*

$$\mathbb{P} [\|\mathbf{X}\|_{\alpha\text{-H\"ol}; \leq \varepsilon; [s, t]} \geq \gamma \mid \mathcal{F}_s] \leq C_{2.6} (t-s) \varepsilon^{-1} (\gamma^{-2} \varepsilon^{1-2\alpha} + 1) \exp \left(\frac{-\gamma^2 (1-2^{-\alpha})^2}{9C_1 \varepsilon^{1-2\alpha}} \right).$$

Proof. Let $\tau_n = \tau_n^{\varepsilon, \gamma, s}$ be defined as in Definition 2.1 with $\tau_0 = s$. Note that (1) implies that for all $c, \gamma > 0$, $t > s$ and $\varepsilon \in (0, t - s]$,

$$\mathbb{P} \left[\exists n \geq 0, \tau_n \leq t, \sup_{u \in [\tau_n, \tau_n + \varepsilon]} d(\mathbf{X}_{\tau_n}, \mathbf{X}_u) > c \mid \mathcal{F}_s \right] \leq \lceil (t-s)/\varepsilon \rceil C_1 \exp \left(\frac{-c^2}{C_1 \varepsilon} \right),$$

so that by Lemma 2.2

$$\mathbb{P} \left[\|\mathbf{X}\|_{\alpha\text{-H\"ol};=\varepsilon;[s,t]} \geq (3c\varepsilon^{-\alpha}) \vee (4\gamma + c\varepsilon^{-\alpha}) \mid \mathcal{F}_s \right] \leq \lceil (t-s)/\varepsilon \rceil C_1 \exp \left(\frac{-c^2}{C_1 \varepsilon} \right).$$

In particular, choosing $c = 2\gamma\varepsilon^\alpha$ yields that for all $\gamma > 0$, $t > s$, and $\varepsilon \in (0, t-s]$,

$$\mathbb{P} \left[\|\mathbf{X}\|_{\alpha\text{-H\"ol};=\varepsilon;[s,t]} \geq 6\gamma \mid \mathcal{F}_s \right] \leq \lceil (t-s)/\varepsilon \rceil C_1 \exp \left(\frac{-(2\gamma)^2}{C_1 \varepsilon^{1-2\alpha}} \right).$$

Hence

$$\mathbb{P} \left[\exists n \geq 0, \|\mathbf{X}\|_{\alpha\text{-H\"ol};=2^{-n}\varepsilon;[s,t]} \geq \gamma \mid \mathcal{F}_s \right] \leq 2C_1(t-s)\varepsilon^{-1} \sum_{n=0}^{\infty} 2^n \exp \left(\frac{-2^{n(1-2\alpha)}\gamma^2}{9C_1\varepsilon^{1-2\alpha}} \right).$$

The conclusion now follows from Lemma 2.3 and the observation that for every $\theta > 0$ there exists C_4 such that for all $K > 0$

$$\sum_{n=0}^{\infty} 2^n \exp(-K2^{\theta n}) \leq C_4(K^{-1} + 1)e^{-K}$$

(which can be seen, for example, by the integral test and the asymptotic behaviour of the incomplete gamma function $\Gamma(p, K)$). \square

Proof of Theorem 2.5. For $\gamma, \varepsilon > 0$ and $s \in [0, T]$, consider the event

$$A_s = \{ \|\mathbf{X}\|_{\alpha\text{-H\"ol};\leq\varepsilon;[s,s+\varepsilon]} < \gamma, \mathbf{X}_{s+\varepsilon} \in B(x, C_2\varepsilon^{1/2}) \}.$$

Applying condition (2) and Lemma 2.6 with $t = s + \varepsilon$, we see that for all $s \in [0, T]$, and $\varepsilon, \gamma > 0$

$$\mathbb{P}[A_s \mid \mathcal{F}_s] \geq c_2 \mathbf{1}\{\mathbf{X}_s \in B(x, C_2\varepsilon^{1/2})\} - C_{2.6}(\gamma^{-2}\varepsilon^{1-2\alpha} + 1) \exp \left(\frac{-\gamma^2(1-2^{-\alpha})^2}{9C_1\varepsilon^{1-2\alpha}} \right).$$

Observe also that Lemma 2.4 (with $r = C_2\varepsilon^{1/2}$) implies that for all $\varepsilon, \gamma > 0$

$$\mathbb{P} \left[\|\mathbf{X}\|_{\alpha\text{-H\"ol};[0,T]} < 2\gamma + 2C_2\varepsilon^{1/2-\alpha} \mid \mathcal{F}_0 \right] \geq \mathbb{P} \left[\bigcap_{k=0}^{\lceil T/\varepsilon \rceil - 1} A_{k\varepsilon} \mid \mathcal{F}_0 \right].$$

It remains to control the final probability on the RHS. We set $\varepsilon = c_1\gamma^{2/(1-2\alpha)}$ (so that $\varepsilon^{1/2-\alpha} \sim \gamma$), where $c_1 > 0$ is sufficiently small (and depends only on $C_1, c_2, C_2, C_{2.6}$ and α) such that

$$\kappa := c_2 - C_{2.6}(c_1^{1-2\alpha} + 1) \exp \left(\frac{-(1-2^{-\alpha})^2}{36C_1c_1^{1-2\alpha}} \right) > 0,$$

so in particular for all $s \in [0, T]$ and $\gamma > 0$,

$$\mathbb{P}[A_s \mid \mathcal{F}_s] \geq \kappa \mathbf{1}\{\mathbf{X}_s \in B(x, C_2\varepsilon^{1/2})\}.$$

Inductively applying conditional expectations, it follows that for all $n \geq 0$

$$\mathbb{P} \left[\bigcap_{k=0}^n A_{k\varepsilon} \mid \mathcal{F}_0 \right] \geq \kappa^{n+1} \mathbf{1}\{\mathbf{X}_0 \in B(x, C_2\varepsilon^{1/2})\}.$$

Taking $n = \lceil T/\varepsilon \rceil - 1$ yields the desired result. \square

2.3. Support theorem for Markovian rough paths. We now turn to the support theorem for Markovian rough paths in α -Hölder topology, which we state in Theorem 2.11 at the end of this section.

Recall the Sobolev path space $W^{1,2} = W^{1,2}([0, T], \mathbb{R}^d)$ and the translation operator $T_h(\mathbf{x})$ defined for $\mathbf{x} \in C^{p\text{-var}}([0, T], G)$, $1 \leq p < N+1$, and $h \in C^{1\text{-var}}([0, T], \mathbb{R}^d)$ (see [FV10, Sec. 1.4.2, 9.4.6]). Let us fix $\alpha \in (0, 1/2)$ and $\Lambda > 0$. Recall further Notation 1.1, in particular the set $\Xi(\Lambda)$.

Proposition 2.7. *Let $h \in W^{1,2}$. There exists a constant $C_{2.7} > 0$, depending only on Λ , $\|h\|_{W^{1,2}}$, α , and T , such that for all $a \in \Xi(\Lambda)$, $x \in G$, and $\gamma > 0$*

$$\mathbb{P}^{a,x} [\|T_h(\mathbf{X})\|_{\alpha\text{-Höl};[0,T]} < \gamma] \geq C_{2.7}^{-1} \exp\left(\frac{-C_{2.7}}{\gamma^{2/(1-2\alpha)}}\right).$$

For the proof, let us fix $h \in W^{1,2}$ and a filtration $(\mathcal{F}_t)_{t \in [0, T]}$ to which \mathbf{X} (and thus $T_h(\mathbf{X})$) is adapted (e.g, the natural filtration generated to \mathbf{X}).

Remark 2.8. If $a(x)$ depends only on the first level $\pi_1(x)$ for all $x \in G$, then $T_h(\mathbf{X})$ is a (non-symmetric, time-inhomogeneous) Markov process. In general, however, $T_h(\mathbf{X})$ is non-Markov. The reason is that, for any fixed $t \in (0, T]$, the sigma-algebra $\sigma(\mathbf{X}_t)$ is not necessarily contained in $\sigma(T_h(\mathbf{X})_t)$, i.e., information on whether $T_h(\mathbf{X})_t \in A$ for Borel subsets $A \subset G$ does not yield full information about \mathbf{X}_t , which is necessary to determine the evolution of \mathbf{X} , and thus of $T_h(\mathbf{X})$.

Recall that the Fernique estimate [FV10, Cor. 16.12] implies that for every stopping time τ and $p > 2$, a.s.

$$(2.3) \quad \mathbb{P} [\|\mathbf{X}\|_{p\text{-var};[\tau, \tau+\varepsilon]} > c \mid \mathcal{F}_\tau] \leq C_F \exp\left(\frac{-c^2}{C_F \varepsilon}\right),$$

where C_F depends only on Λ and p . We now prove two lemmas which demonstrate that the process $T_h(\mathbf{X})$ satisfies conditions (1) and (2).

Lemma 2.9. *There exists a constant $C > 0$, depending only on Λ , such that for all $c, \varepsilon > 0$ satisfying*

$$(2.4) \quad \varepsilon \leq \frac{c^2}{4\|h\|_{W^{1,2}}^2},$$

it holds that for every stopping time τ , a.s.

$$\mathbb{P} \left[\sup_{t \in [\tau, \tau+\varepsilon]} d(T_h(\mathbf{X})_\tau, T_h(\mathbf{X})_t) > c \mid \mathcal{F}_\tau \right] \leq C \exp\left(\frac{-c^2}{4C\varepsilon}\right).$$

Proof. Suppose $c, \varepsilon > 0$ satisfy (2.4). Using that $\|h\|_{1\text{-var};[s, s+\varepsilon]} \leq \varepsilon^{1/2} \|h\|_{W^{1,2};[s, s+\varepsilon]}$, we have $\|h\|_{1\text{-var};[s, s+\varepsilon]} \leq c/2$. Fix now any $2 < p < N+1$. Observe that (see [FV10, Thm. 9.33])

$$\begin{aligned} \sup_{t \in [s, s+\varepsilon]} d(T_h(\mathbf{X})_s, T_h(\mathbf{X})_t) &\leq \|T_h(\mathbf{X})\|_{p\text{-var};[s, s+\varepsilon]} \\ &\leq C_1 (\|\mathbf{X}\|_{p\text{-var};[s, s+\varepsilon]} + \|h\|_{1\text{-var};[s, s+\varepsilon]}), \end{aligned}$$

from which the conclusion follows by the Fernique estimate (2.3). \square

Lemma 2.10. *For all $C \geq C_0(\Lambda, \|h\|_{W^{1,2}}) > 0$, there exists $c = c(C, \Lambda, \|h\|_{W^{1,2}}) > 0$ such that for all $x \in G$, $s \in [0, T]$, and $\varepsilon \in (0, T-s]$, a.s.*

$$\mathbb{P}[T_h(\mathbf{X})_{s+\varepsilon} \in B(x, C\varepsilon^{1/2}) \mid \mathcal{F}_s] \geq c \mathbf{1}\{T_h(\mathbf{X})_s \in B(x, C\varepsilon^{1/2})\}.$$

Proof. We use the shorthand notation $\mathbf{Y} = T_h(\mathbf{X})$. For every $x, y \in G$, consider a geodesic $\gamma^{y,x} : [0, 1] \rightarrow G$ with $\gamma_0^{y,x} = y$ and $\gamma_1^{y,x} = x$ parametrised at unit speed. Let $z(y, x) := \gamma_{1/2}^{y,x}$ denote its midpoint. For any $x \in G$, observe that

$$\begin{aligned} d(\mathbf{Y}_{s+\varepsilon}, x) &\leq d(\mathbf{Y}_{s+\varepsilon}, z(\mathbf{Y}_s, x)) + d(z(\mathbf{Y}_s, x), x) \\ &\leq d(\mathbf{Y}_{s,s+\varepsilon}, \mathbf{X}_{s,s+\varepsilon}) + d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_s^{-1}z(\mathbf{Y}_s, x)) + d(z(\mathbf{Y}_s, x), x). \end{aligned}$$

If $\mathbf{Y}_s \in B(x, r)$, then evidently $d(z(\mathbf{Y}_s, x), x) \leq r/2$. Moreover, since G is a homogeneous group and due to our normalisation of λ , it holds that $\lambda(B(x, r)) = r^Q$ for all $r \geq 0$ and $x \in G$, where $Q \geq 1$ is the homogeneous dimension of G . Recall also the lower bound on the heat kernel [FV10, Thm. 16.11]

$$p(\varepsilon, x, y) \geq C_l^{-1} \varepsilon^{-Q/2} \exp\left(\frac{-C_l d(x, y)^2}{\varepsilon}\right), \quad \forall x, y \in G, \quad \forall \varepsilon > 0,$$

where $C_l > 0$ depends only on Λ . It follows that there exists $C_1 > 0$, depending only on Λ , such that, for any $r, \varepsilon > 0$ and $y \in B(1_G, r/2)$,

$$\begin{aligned} \mathbb{P}[d(\mathbf{X}_{s,s+\varepsilon}, y) < r/4] &\geq \lambda(B(y, r/4)) C_l^{-1} \varepsilon^{-Q/2} \exp\left(\frac{-C_l r^2}{\varepsilon}\right) \\ &\geq \frac{C_1^{-1} r^Q}{\varepsilon^{Q/2}} \exp\left(\frac{-C_1 r^2}{\varepsilon}\right). \end{aligned}$$

Note that if $\mathbf{Y}_s \in B(x, r)$, then necessarily $\mathbf{Y}_s^{-1}z(\mathbf{Y}_s, x) \in B(1_G, r/2)$, so we obtain for all $x \in G$, $r, \varepsilon > 0$ and $s \in [0, T]$

$$\mathbb{P}[d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_s^{-1}z(\mathbf{Y}_s, x)) < r/4 \mid \mathcal{F}_s] \geq \frac{C_1^{-1} r^Q}{\varepsilon^{Q/2}} \exp\left(\frac{-C_1 r^2}{\varepsilon}\right) \mathbf{1}\{\mathbf{Y}_s \in B(x, r)\}.$$

Finally, by standard rough paths estimates (using that $T_h(\mathbf{X})_{s,t}$ is equal to $\mathbf{X}_{s,t}$ plus a combination of cross-integrals of \mathbf{X} and h over $[s, t]$) we have

$$\begin{aligned} d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_{s,s+\varepsilon}) &\leq C_2 \max_{i \in \{1, \dots, N\}} \left(\sum_{k=1}^i \|h\|_{1\text{-var};[s,s+\varepsilon]}^k \|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]}^{i-k} \right)^{1/i} \\ &\leq C_2 \max_{i \in \{1, \dots, N\}} \left(\sum_{k=1}^i \varepsilon^{k/2} \|h\|_{W^{1,2};[s,s+\varepsilon]}^k \|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]}^{i-k} \right)^{1/i}. \end{aligned}$$

Hence, if $\|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]} \leq R\varepsilon^{1/2}$, then for some $C_3 > 0$ depending only on G

$$d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_{s,s+\varepsilon}) \leq C_3 \varepsilon^{1/2} (\|h\|_{W^{1,2};[s,s+\varepsilon]}^{1/N} + \|h\|_{W^{1,2};[s,s+\varepsilon]})(1 + R^{(N-1)/N}).$$

We now let $r = C\varepsilon^{1/2}$. It follows that if C and R satisfy

$$(2.5) \quad C \geq 4C_3 (\|h\|_{W^{1,2};[s,s+\varepsilon]}^{1/N} + \|h\|_{W^{1,2};[s,s+\varepsilon]})(1 + R^{(N-1)/N}),$$

then by the Fernique estimate (2.3), for any $2 < p < N + 1$,

$$\begin{aligned} \mathbb{P}[d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_{s,s+\varepsilon}) > C\varepsilon^{1/2}/4 \mid \mathcal{F}_s] &\leq \mathbb{P}[\|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]} > R\varepsilon^{1/2} \mid \mathcal{F}_s] \\ &\leq C_F \exp\left(\frac{-R^2}{C_F}\right). \end{aligned}$$

It follows that if C and R furthermore satisfy

$$(2.6) \quad c := C_1^{-1} C^Q \exp(-C_1 C^2) - C_F \exp\left(\frac{-R^2}{C_F}\right) > 0,$$

then we obtain

$$\mathbb{P}[d(\mathbf{Y}_{s+\varepsilon}, x) < C\varepsilon^{1/2} \mid \mathcal{F}_s] \geq c\mathbf{1}\{\mathbf{Y}_s \in B(x, C\varepsilon^{1/2})\}.$$

We now observe that due to the factor $R^{(N-1)/N}$ in (2.5) above, there exists $C_0 > 0$, depending only on $\|h\|_{W^{1,2}}$ and Λ , such that for every $C \geq C_0$, we can find $R > 0$ for which (2.5) and (2.6) are satisfied. \square

Proof of Proposition 2.7. By Theorem 2.5, it suffices to check that $T_h(\mathbf{X})$ satisfies conditions (1) and (2) with constants C_1, c_2, C_2 only depending on Λ and $\|h\|_{W^{1,2}}$. However this follows directly from Lemmas 2.9 and 2.10. \square

Theorem 2.11. *Let $\gamma, R > 0$. It holds that*

$$(2.7) \quad \inf_{x \in G} \inf_{a \in \Xi(\Lambda)} \inf_{\|h\|_{W^{1,2}} \leq R} \mathbb{P}^{a,x} [d_{\alpha\text{-H\"ol};[0,T]}(\mathbf{X}, S_N(h)) < \gamma] > 0,$$

where $d_{\alpha\text{-H\"ol};[0,T]}$ denotes the (homogeneous) α -H\"older metric and $S_N(h)$ is the level- N lift of h . In particular, the support of $\mathbf{X}^{a,x}$ in α -H\"older topology is precisely the closure in $C^{\alpha\text{-H\"ol}}([0, T], G)$ of $\{xS_N(h) \mid h \in W^{1,2}\}$.

Proof. By uniform continuity of the map $(\mathbf{x}, h) \mapsto T_h(\mathbf{x})$ on bounded sets [FV10, Cor. 9.35], and the fact that $T_h T_{-h}(\mathbf{x}) = \mathbf{x}$ and $T_h(0) = S_N(h)$, there exists $\delta = \delta(\gamma, R) > 0$ such that for all $h \in W^{1,2}$ with $\|h\|_{W^{1,2}} \leq R$

$$\|T_{-h}(\mathbf{x})\|_{\alpha\text{-H\"ol};[0,T]} < \delta(\gamma, R) \Rightarrow d_{\alpha\text{-H\"ol}}(\mathbf{x}, S_N(h)) < \gamma.$$

The bound (2.7) then follows from Proposition 2.7. As a consequence, we see that the support of $\mathbf{X}^{a,x}$ contains the closure of $\{xS_N(h) \mid h \in W^{1,2}\}$. The reverse inclusion follows from the fact that $\mathbf{X}^{a,x}$ is a.s. a geometric α -H\"older rough path, and is therefore the limit in the $d_{\alpha\text{-H\"ol};[0,T]}$ metric of lifts of smooth paths. \square

Remark 2.12. The main difference with the approach taken in [FV08, Thm. 50] and [FV10, Thm 16.33] to prove a bound of the form $\mathbb{P}[d_{\alpha\text{-H\"ol}}(\mathbf{X}, S_N(h)) < \gamma] > 0$ (with $\alpha \in [0, 1/6]$ and $\alpha \in [0, 1/4]$ respectively) is that we do not rely on a support theorem in the uniform topology. As a consequence, our analysis is more delicate but does not lose any power at each step, which allows us to push to the sharp H\"older exponent range $\alpha \in [0, 1/2]$.

Note also that [FV10, Thm 16.39] and [FV08, Cor. 46] give this bound for $h \equiv 0$ with the sharp range $\alpha \in [0, 1/2]$. The proof therein relies crucially on lower and upper bounds on the probability that \mathbf{X} stays in small balls, namely $\mathbb{P}^{a,x}[\|\mathbf{X}\|_{0;[0,t]} < \gamma] \asymp e^{-\lambda(\gamma)t\gamma^2}$ with $0 < \lambda_{\min} \leq \lambda(\gamma) \leq \lambda_{\max} < \infty$, which yields a version of Lemma 2.6 for the untranslated process $\mathbf{X}^{a,x}$ conditioned to stay in a small ball around x . This argument is rather sensitive to the fact that for each fixed $\gamma > 0$ the same quantity $\lambda(\gamma)$ appears in the lower and upper bounds; this is not true for the translated process $T_h(\mathbf{X})$, which is the reason for our different strategy.

3. DENSITY THEOREM

3.1. Semi-Dirichlet forms associated with H\"ormander vector fields. In this subsection, let \mathcal{O} be a smooth manifold and $W = (W_1, \dots, W_d)$ a collection of smooth vector fields on \mathcal{O} . For $z \in \mathcal{O}$, let $\text{Lie}_z W$ denote the subspace of $T_z \mathcal{O}$ spanned by the vector fields (W_1, \dots, W_d) and all their commutators at z . We say

that W satisfies Hörmander's condition on \mathcal{O} if $\text{Lie}_z W = T_z \mathcal{O}$ for every $z \in \mathcal{O}$, in which case we call W a collection of Hörmander vector fields.

Fix a collection $W = (W_1, \dots, W_d)$ of Hörmander vector fields on \mathcal{O} and $U \subset \mathcal{O}$ an open subset with compact closure. Consider a bounded measurable function a on U taking values in (not necessarily symmetric) $d \times d$ matrices such that for some $\Lambda \geq 1$

$$(3.1) \quad \Lambda^{-1}|\xi|^2 \leq \langle \xi, a(z)\xi \rangle, \quad \forall \xi \in \mathbb{R}^d, \quad \forall z \in U.$$

Let μ be a smooth measure on \mathcal{O} and define the bilinear map

$$\begin{aligned} \mathcal{E} : C_c^\infty(U) \times C_c^\infty(U) &\rightarrow \mathbb{R} \\ \mathcal{E} : (f, g) &\mapsto - \sum_{i,j=1}^d \int_U a^{i,j}(z) (W_i f)(z) (W_j^* g)(z) \mu(dz), \end{aligned}$$

where $W_j^* = -W_j - \text{div}_\mu W_j$ is the formal adjoint of W_j with respect to μ . In the following lemma, the L^p norm $\|\cdot\|_p$ for $p \in [1, \infty]$ is assumed to be on $L^p(U, \mu)$. For background concerning (non-symmetric, semi-)Dirichlet forms, we refer to [Osh13].

Lemma 3.1. *The bilinear form \mathcal{E} is closable in $L^2(U, \mu)$, lower bounded, and satisfies the sector condition. Denote by P_t the associated (strongly continuous) semi-group on $L^2(U, \mu)$. Suppose further that P_t is sub-Markov (so that the closed extension of \mathcal{E} is a lower-bounded semi-Dirichlet form) and maps $C_b(U)$ into itself. Then there exists $\nu > 2$ and $b > 0$ such that for every $x \in U$ and $t > 0$ there exists $p_t(x, \cdot) \in L^2(U, \mu)$ with $\|p_t(x, \cdot)\|_2 \leq bt^{-\nu/2}$ such that for all $f \in L^2(U, \mu)$*

$$P_t f(x) = \int_U p_t(x, y) f(y) \mu(dy).$$

The proof of Lemma 3.1 is based on the sub-Riemannian Sobolev inequality combined with a classical argument of Nash [Nas58]. We believe this result should be standard, but as we were unable to find a sufficiently similar form in the literature, we prefer to give a proof in Appendix A (see [SCS91, Stu95] for closely related results in the case that \mathcal{E} is symmetric or positive semi-definite).

Note also that in the sequel, namely in the proof of Theorem 3.4, we will only require the fact from Lemma 3.1 that the kernel p_t exists. The bound on $\|p_t(x, \cdot)\|_2$ is merely a free consequence of the proof of its existence.

3.2. Density for RDEs. We now specialise to the setting of Markovian rough paths. Recall Notation 1.1 and consider the RDE

$$(3.2) \quad d\mathbf{Y}_t = V(\mathbf{Y}_t) d\mathbf{X}_t, \quad \mathbf{Y}_0 = y_0 \in \mathbb{R}^e,$$

for smooth vector fields $V = (V_1, \dots, V_d)$ on \mathbb{R}^e . We suppose also that V are Lip^2 so that (3.2) admits a unique solution. We fix also the starting point $\mathbf{X}_0 = x_0 \in G$ of \mathbf{X} .

For the reader's convenience, we recall the Nagano–Sussmann orbit theorem (see, e.g., [AS04, Chpt. 5]).

Theorem 3.2 (Orbit theorem, Nagano–Sussmann). *Let W be a set of complete smooth vector fields on a smooth manifold M . Let \mathcal{O} denote the orbit of W through a point $z_0 \in M$. Then \mathcal{O} is a connected immersed submanifold of M . Furthermore, for any $z \in \mathcal{O}$,*

$$T_z \mathcal{O} = \text{span} \left\{ d(P^{-1})_{P(z)} w(P(z)) \mid P \in \mathcal{P}, w \in W \right\},$$

where

$$\mathcal{P} = \{e^{t_1 w_1} \circ \dots \circ e^{t_k w_k} \mid t_i \in \mathbb{R}, w_i \in W, k \geq 1\} \subset \text{Diff } M.$$

A particularly useful consequence of the orbit theorem is the following.

Corollary 3.3. *Let notation be as in Theorem 3.2. It holds that $\text{Lie}_z W \subseteq T_z \mathcal{O}$ for all $z \in \mathcal{O}$. Furthermore, $\text{Lie}_z W = T_z \mathcal{O}$ for all $z \in \mathcal{O}$ if and only if $\dim \text{Lie}_z W$ is constant in z .*

Proof. The fact that $\text{Lie}_z W \subseteq T_z \mathcal{O}$ and the “only if” implication are obvious. For the “if” implication, suppose $\dim \text{Lie}_z W$ is constant in $z \in \mathcal{O}$. Then $\text{Lie } W$ defines a distribution on \mathcal{O} (a subbundle of the tangent bundle), so the Frobenius theorem implies that $\text{Lie } W$ arises from a regular foliation of \mathcal{O} . However, each leaf of this foliation is itself an orbit of W . Therefore the foliation contains only one leaf, namely \mathcal{O} , which concludes the proof. \square

Consider the manifold $G \times \mathbb{R}^e$. We canonically identify the tangent space $T_{(x,y)}(G \times \mathbb{R}^e)$ with $T_x G \oplus T_y \mathbb{R}^e$ and define smooth vector fields on $G \times \mathbb{R}^e$ by $W_i = U_i + V_i$. Let $z_0 = (x_0, y_0) \in G \times \mathbb{R}^e$ and denote by $\mathcal{O} = \mathcal{O}_{z_0}$ the orbit of z_0 under the collection $W = (W_1, \dots, W_d)$.

Denote the couple $\mathbf{Z}_t = (\mathbf{X}_t, \mathbf{Y}_t)$ which is a Markov process on $G \times \mathbb{R}^e$. One can readily show that a.s. $\mathbf{Z}_t^{z_0} \in \mathcal{O}$ for all $t > 0$ (e.g., by approximating each sample path of \mathbf{X} in p -variation for some $p > 2$ by piecewise geodesic paths).

Theorem 3.4. *Suppose W satisfies Hörmander’s condition on \mathcal{O} , i.e., $\text{Lie}_z W = T_z \mathcal{O}$ for all $z \in \mathcal{O}$. Then for all $t > 0$, $\mathbf{Z}_t^{z_0}$ admits a density with respect to any smooth measure on \mathcal{O} .*

The proof of Theorem 3.4 will be given at the end of this section. We first state several remarks and a consequence of the theorem.

Remark 3.5. Note that from Notation 1.1 we always consider $G = G^N(\mathbb{R}^d)$ with $N \geq 2$. However, in the special case that $a(x)$ depends only on the first level $\pi_1(x)$ for all $x \in G^N(\mathbb{R}^d)$, the identical statement in Theorem 3.4 holds for the process $\mathbf{Z}_t = (\pi_1(\mathbf{X}_t), \mathbf{Y}_t) \in \mathbb{R}^d \times \mathbb{R}^e$ (the conditions change by substituting G by \mathbb{R}^d everywhere). The reason for this is that Lemma 3.13 below can be readily adjusted to give analogous infinitesimal behaviour of the process \mathbf{Z}_t (now taking values in $\mathcal{O} \subseteq \mathbb{R}^d \times \mathbb{R}^e$), after which the proof of the theorem carries through without change.

For a statement of the density of \mathbf{Y}_t itself, let $\mathcal{O}' \subseteq \mathbb{R}^e$ denote the orbit of $y_0 \in \mathbb{R}^e$ under V .

Lemma 3.6. *Suppose $\mathbf{Z}_t^{z_0}$ admits a density with respect to a smooth measure on \mathcal{O} . Then \mathbf{Y}_t admits a density with respect to any smooth measure on \mathcal{O}' .*

Proof. By the description of the tangent space $T_z \mathcal{O}$ in Theorem 3.2, it holds that the projection $p_2 : \mathcal{O} \rightarrow \mathcal{O}', (x, y) \mapsto y$, is a (surjective) submersion (in fact a smooth fibre bundle) from \mathcal{O} to \mathcal{O}' . The conclusion follows from the fact that pre-images of null-sets under submersions are null-sets for smooth measures. \square

Moreover, the condition in Theorem 3.4 may be restated in terms of just the driving vector fields $V = (V_1, \dots, V_d)$ as follows.

Lemma 3.7. *For a multi-index $I = (i_1, \dots, i_k) \in \{1, \dots, d\}^k$ of length $|I| = k$, denote by $V_{[I]}$ the vector field $[[\dots [V_{i_1}, V_{i_2}], \dots], V_{i_k}]$. It holds that W satisfies Hörmander's condition on \mathcal{O} if and only if*

$$(3.3) \quad \dim \text{span}\{V_{[I]}(y) : |I| > N\} \subseteq T_y \mathbb{R}^e \text{ is constant in } y \in \mathcal{O}'.$$

Proof. Since the vector fields U_1, \dots, U_d are freely step- N nilpotent and generate the tangent space of G , observe that

$$(3.4) \quad \dim \text{Lie}_{(x,y)} W = \dim G + \dim \text{span}\{V_{[I]}(y) : |I| > N\}, \quad \forall (x, y) \in G \times \mathbb{R}^e.$$

Suppose W satisfies Hörmander's condition on \mathcal{O} . Then $\dim \text{Lie}_z W$ is constant in $z \in \mathcal{O}$, and by (3.4) it follows that (3.3) holds. Conversely, suppose (3.3) holds. It now follows from (3.4) that $\dim \text{Lie}_z W$ is constant in $z \in \mathcal{O}$, and thus W satisfies Hörmander's condition on \mathcal{O} by Corollary 3.3. \square

Combining Theorem 3.4 with Lemmas 3.6 and 3.7, we obtain the following corollary.

Corollary 3.8. *Suppose condition (3.3) holds. Then for all $t > 0$, the RDE solution \mathbf{Y}_t admits a density with respect to any smooth measure on \mathcal{O}' .*

Remark 3.9. Note that $\mathcal{O}' = \mathbb{R}^e$ whenever V satisfies Hörmander's condition on \mathbb{R}^e , in which case every smooth measure is equivalent to the Lebesgue measure.

Remark 3.10. Following Remark 3.5, in the case that $a(x)$ depends only on the first level $\pi_1(x)$, we are able to take $N = 1$ in (3.3) when applying Corollary 3.8.

Remark 3.11. Note that while (3.3) (for any $N \geq 0$) implies that V satisfies Hörmander's condition on \mathcal{O}' , the reverse implication is clearly not true. In particular, we do not know if it is sufficient for V to only satisfy Hörmander's condition on \mathcal{O}' in order for \mathbf{Y}_t to admit a density on \mathcal{O}' . The difficulty of course is that unless (3.3) is satisfied, the couple $(\mathbf{X}_t, \mathbf{Y}_t)$ will in general not admit a density in \mathcal{O} , whereby our method of proof breaks down.

For the proof of Theorem 3.4, we first recall for the reader's convenience the infinitesimal behaviour of the coordinate projections of \mathbf{X}^a . As before, let λ denote the Haar measure on G .

Lemma 3.12. *Let $g \in C_c^\infty(G)$. Then for all $k, l \in \{1, \dots, d\}$*

$$\begin{aligned} \lim_{t \rightarrow 0} t^{-1} \langle g, \mathbb{E}^{a, \cdot} [\mathbf{X}_{0,t}^k] \rangle_{L^2(G, \lambda)} &= - \sum_{j=1}^d \int_G a^{k,j}(x) U_j g(x) \lambda(dx), \\ \lim_{t \rightarrow 0} t^{-1} \langle g, \mathbb{E}^{a, \cdot} [\mathbf{X}_{0,t}^k \mathbf{X}_{0,t}^l] \rangle_{L^2(G, \lambda)} &= 2 \int_G a^{k,l}(x) g(x) \lambda(dx), \\ \lim_{t \rightarrow 0} t^{-1} \langle g, \mathbb{E}^{a, \cdot} [\mathbf{X}_{0,t}^{k,l}] \rangle_{L^2(G, \lambda)} &= 0. \end{aligned}$$

Proof. This is [FV08, Lem. 27] extended *mutatis mutandis* to the general case $G^N(\mathbb{R}^d)$, $N \geq 1$, cf. [FV10, Prop. 16.20]. \square

Lemma 3.13. *Let $U \subset \mathcal{O}$ be an open subset with compact closure. Consider the (sub-Markov) semi-group P_t^U of \mathbf{Z}_t killed upon exiting U , defined for all bounded measurable $f : U \rightarrow \mathbb{R}$ by*

$$P_t^U f(z) = \mathbb{E}^z [f(\mathbf{Z}_t) \mathbf{1}\{\mathbf{Z}_s \in U, \forall s \in [0, t]\}].$$

Then P_t^U maps $C_b(U)$ into itself, and for any smooth measure μ on \mathcal{O} it holds that for all $f, g \in C_c^\infty(U)$

$$(3.5) \quad \lim_{t \rightarrow 0} t^{-1} \langle P_t^U f - f, g \rangle_{L^2(U, \mu)} = \sum_{i,j=1}^d \int_U a^{i,j}(p_1(z)) (W_i f)(z) (W_j^* g)(z) \mu(dz),$$

where $p_1 : \mathcal{O} \rightarrow G$ is the projection $(x, y) \mapsto x$ and $W_j^* = -W_j - \operatorname{div}_\mu(W_j)$ is the adjoint of W_j in $L^2(U, \mu)$.

Proof. To show that P_t^U maps $C_b(U)$ into itself, let $f \in C_b(U)$. As $z_n = (x_n, y_n) \rightarrow z = (x, y)$ in U , it holds in particular that $x_n \rightarrow x$ in G . It follows that $\mathbf{X}^{a, x_n} \xrightarrow{\mathcal{D}} \mathbf{X}^{a, x}$ in α -Hölder topology for any $\alpha \in [0, 1/2)$ [FV10, Thm. 16.28], and we readily obtain that $P_t^U f(z_n) \rightarrow P_t^U f(z)$. Hence $P_t^U f \in C_b(U)$, so indeed P_t^U maps $C_b(U)$ into itself.

It remains to verify (3.5). Note that for every $z \in U$ the probability that \mathbf{Z}^z leaves U in $[0, t]$ is bounded above by $C^{-1} \exp(-Ct^{-1})$ for some $C = C(z, U, \Lambda) > 0$ (see, e.g., the Fernique estimate (2.3)). It follows by a localisation argument and the stochastic Taylor expansion (e.g., [FV08, Lem. 26]), that

$$(3.6) \quad \begin{aligned} \lim_{t \rightarrow 0} t^{-1} \langle P_t^U f - f, g \rangle_{L^2(U, \mu)} &= \lim_{t \rightarrow 0} t^{-1} \int_U \left(\sum_{i=1}^d W_i f(z) \mathbb{E}^x [\mathbf{X}_{0,t}^i] \right. \\ &\quad + \sum_{i,j=1}^d \frac{1}{2} W_i W_j f(z) \mathbb{E}^x [\mathbf{X}_{0,t}^i \mathbf{X}_{0,t}^{x;j}] \\ &\quad \left. + \sum_{i,j=1}^d \frac{1}{2} [W_i, W_j] f(z) \mathbb{E}^x [\mathbf{X}_{0,t}^{i,j}] \right) g(z) \mu(dz). \end{aligned}$$

Since $p_1 : \mathcal{O} \rightarrow G$ is a (surjective) submersion (in fact a smooth fibre bundle), by integrating over the fibres (e.g., [GS77, p. 307]) we can associate to any $v \in C_c^\infty(U)$ a function $\hat{v} \in C_c^\infty(G)$ such that for any bounded measurable $h : G \rightarrow \mathbb{R}$

$$\int_U (h \circ p_1)(z) v(z) \mu(dz) = \int_G h(x) \hat{v}(x) \lambda(dx).$$

In particular, setting $v_i := (W_i f)g$, $v_{i,j} := (W_i W_j f)g$ and $w_{i,j} := ([W_i, W_j] f)g$, we can apply Lemma 3.12 to obtain that (3.6) equals

$$(3.7) \quad \sum_{i,j=1}^d \int_G [-a^{i,j}(x) (U_j \hat{v}_i)(x) + a^{i,j}(x) \hat{v}_{i,j}(x)] \lambda(dx).$$

It remains to show that (3.7) agrees with the RHS of (3.5). To this end, we may assume by a limiting procedure that a is smooth, and note that the same argument as in [FV08, p. 503] applies *mutatis mutandis* to our current setting. \square

Proof of Theorem 3.4. Consider an increasing sequence of relatively compact open sets $(U_n)_{n \geq 1}$ such that $\cup_{n \geq 1} U_n = \mathcal{O}$. By Lemma 3.13, we can apply Lemma 3.1 to conclude that for every $x \in \mathcal{O}$ and $n \geq 1$ such that $x \in U_n$, there exists a non-negative kernel $p_t^n(x, \cdot) \in L^2(U_n, \mu)$ such that $P_t^{U_n} f(x) = \langle p_t^n(x, \cdot), f \rangle_{L^2(U_n, \mu)}$ for all $f \in C_b(U_n)$. Moreover, by definition of $P_t^{U_n}$, the sequence $p_t^n(x, \cdot)$ is increasing in n and satisfies $\|p_t^n(x, \cdot)\|_{L^1(U_n, \mu)} \leq 1$. Hence the limit $p_t(x, \cdot) := \lim_{n \rightarrow \infty} p_t^n(x, \cdot)$

is almost everywhere finite and gives precisely the transition kernel of the Markov process \mathbf{Z}_t in \mathcal{O} with respect to μ . \square

Remark 3.14. The pre-compact subsets U_n were considered in the proof only to obtain existence of p_t^n from Lemma 3.1 for each $n \geq 1$. We could have avoided considering such a compact exhaustion by formulating Lemma 3.1 without a pre-compactness assumption on U (however, at least without extra assumptions, the proof of such a formulation itself would seem to require a compact exhaustion).

APPENDIX A. PROOF OF LEMMA 3.1

We follow the notation from Section 3.1. For $f \in C_c^\infty(U)$ denote

$$\|Wf\|_2^2 := \sum_{i=1}^d \|W_i f\|_2^2,$$

and for $\alpha > 0$

$$\mathcal{E}_\alpha(f, f) := \mathcal{E}(f, f) + \alpha \|f\|_2^2.$$

Lemma A.1. (1) For every $\varepsilon < \Lambda^{-1}$, there exists $\alpha > 0$, depending only on ε , Λ , $\|a\|_\infty$ and $\sum_{i=1}^d \|\operatorname{div}_\mu W_i\|_\infty$, such that for all $f, g \in C_c^\infty(U)$

$$(A.1) \quad \mathcal{E}_\alpha(f, f) \geq \varepsilon \|Wf\|_2^2.$$

(2) There exist $\beta > 0$, depending only on $\|a\|_\infty$ and $\sum_{i=1}^d \|\operatorname{div}_\mu W_i\|_\infty$, such that

$$(A.2) \quad |\mathcal{E}(f, g)| \leq \beta \|Wf\|_2 (\|Wg\|_2 + \|g\|_2).$$

Proof. By the Cauchy-Schwartz inequality and (3.1), for some $C_1, \alpha > 0$

$$\begin{aligned} \mathcal{E}(f, f) &= \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) W_j f(z) \mu(dz) + \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) \operatorname{div}_\mu W_j(z) f(z) \mu(dz) \\ &\geq \sum_{i=1}^d \int_U \Lambda^{-1} |W_i f(z)|^2 \mu(dz) - \sum_{i,j=1}^d \|\operatorname{div}_\mu W_j\|_\infty \|a^{i,j}\|_\infty \|W_i f\|_2 \|f\|_2 \\ &\geq \Lambda^{-1} \|Wf\|_2^2 - C_1 \|Wf\|_2 \|f\|_2 \\ &\geq \varepsilon \|Wf\|_2^2 - \alpha \|f\|_2^2, \end{aligned}$$

which implies (A.1). On the other hand, by Cauchy-Schwartz, for some $C_2, C_3 > 0$

$$\left| \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) W_j g(z) dz \right| \leq C_2 \|Wf\|_2 \|Wg\|_2,$$

and

$$\left| \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) \operatorname{div}_\mu W_j(z) g(z) dz \right| \leq C_3 \|Wf\|_2 \|g\|_2,$$

from which we obtain (A.2). \square

Since W satisfies Hörmander's condition on \mathcal{O} , recall that for every $x \in \mathcal{O}$ there exist constants $\nu_x > 2$, $C_x > 0$, and a neighbourhood U_x of x with $\mu(U_x) < \infty$ such that for all $f \in C_c^\infty(U_x)$ (see, e.g., [Stu95, p. 296])

$$\left(\int_{U_x} |f|^{2\nu_x/(\nu_x-2)} d\mu \right)^{(\nu_x-2)/\nu_x} \leq C_x \int_{U_x} \left(\sum_{i=1}^d |W_i f|^2 + |f|^2 \right) d\mu.$$

Since U is pre-compact, it is routine to patch together such inequalities using a partition of unity and apply interpolation to arrive at the following Sobolev inequality.

Lemma A.2 (Sobolev inequality). *There exist constants $\nu > 2$ and $C, R > 0$ such that for all $f \in C_c^\infty(U)$ with $\|f\|_1 \leq 1$ and $\|f\|_2 > R$*

$$\|f\|_{2\nu/(\nu-2)}^2 \leq C \|Wf\|_2^2.$$

Fix $\varepsilon < \Lambda^{-1}$ and $\alpha > 0$ such that (A.1) holds. Let \mathcal{F} be the closure of $C_c^\infty(U)$ under $\|\cdot\|_{\mathcal{F}} := \mathcal{E}_\alpha(\cdot, \cdot)^{1/2}$.

Corollary A.3 (Nash inequality). *Let $\nu > 2$ and $R > 0$ be the same as in Lemma A.2. There exists $c > 0$ such that for all $f \in \mathcal{F}$ with $\|f\|_1 \leq 1$ and $\|f\|_2 > R$, it holds that*

$$(A.3) \quad \mathcal{E}(f, f) \geq c \|f\|_2^{2+4/\nu}.$$

Proof. Consider first $f \in C_c^\infty(U)$. The Sobolev inequality (Lemma A.2), along with Hölder's inequality, implies that

$$\|f\|_2^{2+4/\nu} \leq C \|f\|_1^{4/\nu} \|Wf\|_2^2,$$

from which the conclusion follows first for all $f \in C_c^\infty(U)$ by (A.1), and then for general $f \in \mathcal{F}$ by an approximation argument. \square

Proof of Lemma 3.1. The desired properties of \mathcal{E} all follow from (A.1) and (A.2) and the fact that each W_i is a closable operator defined on $C_c^\infty(U) \subset L^2(U, \mu)$.

Denote by A the generator of the associated adjoint semi-group P_t^* in $L^2(U, \mu)$ with domain $D(A)$. Consider $f \in D(A)$ with $\|f\|_1 \leq 1$ and set $u_t = P_t^* f$. Since P_t is sub-Markov, we have $\|u_t\|_1 \leq 1$, so by Corollary A.3, whenever $\|u_t\|_2 > R$,

$$\frac{d}{dt} \|u_t\|_2^2 = \lim_{h \rightarrow 0} \frac{\|P_h^* u_t\|_2^2 - \|u_t\|_2^2}{h} = -2\mathcal{E}(u_t, u_t) \leq -2c \|u_t\|_2^{2+4/\nu},$$

from which it follows that there exists $b > 0$ such that $\|P_t^* f\|_2 \leq bt^{-\nu/2}$.

To complete the proof, it remains only to apply an approximation of the Dirac delta $\langle \phi, f_n \rangle \rightarrow \langle \phi, \delta_x \rangle = \phi(x)$ for all $\phi \in C_b(U)$, with $f_n \in D(A)$ and $\|f_1\| \leq 1$, and use the fact that $\sup_n \|P_t^* f_n\|_2 \leq bt^{-\nu/2}$ and that P_t preserves $C_b(U)$. \square

REFERENCES

- [AS04] A. A. AGRACHEV and Y. L. SACHKOV. *Control theory from the geometric viewpoint*, vol. 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004, xiv+412. Control Theory and Optimization, II. doi:10.1007/978-3-662-06404-7.
- [BH07] F. BAUDOIN and M. HAIRER. A version of Hörmander's theorem for the fractional Brownian motion. *Probab. Theory Related Fields* **139**, no. 3-4, (2007), 373-395. doi:10.1007/s00440-006-0035-0.
- [BHL02] R. F. BASS, B. M. HAMBLY, and T. J. LYONS. Extending the Wong-Zakai theorem to reversible Markov processes. *J. Eur. Math. Soc. (JEMS)* **4**, no. 3, (2002), 237-269. doi:10.1007/s100970200040.

- [CF10] T. CASS and P. FRIZ. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)* **171**, no. 3, (2010), 2115–2141. doi:10.4007/annals.2010.171.2115.
- [Che18] I. CHEVYREV. Random walks and Lévy processes as rough paths. *Probab. Theory Related Fields* **170**, no. 3-4, (2018), 891–932. doi:10.1007/s00440-017-0781-1.
- [CHLT15] T. CASS, M. HAIRER, C. LITTERER, and S. TINDEL. Smoothness of the density for solutions to Gaussian rough differential equations. *Ann. Probab.* **43**, no. 1, (2015), 188–239. doi:10.1214/13-AOP896.
- [CL16] I. CHEVYREV and T. LYONS. Characteristic functions of measures on geometric rough paths. *Ann. Probab.* **44**, no. 6, (2016), 4049–4082. doi:10.1214/15-AOP1068.
- [CO17] T. CASS and M. OGRODNIK. Tail estimates for Markovian rough paths. *Ann. Probab.* **45**, no. 4, (2017), 2477–2504.
- [FGGR16] P. K. FRIZ, B. GESS, A. GULISASHVILI, and S. RIEDEL. The Jain–Monrad criterion for rough paths and applications to random Fourier series and non-Markovian Hörmander theory. *Ann. Probab.* **44**, no. 1, (2016), 684–738. doi:10.1214/14-AOP986.
- [Fit97] P. J. FITZSIMMONS. Absolute continuity of symmetric diffusions. *Ann. Probab.* **25**, no. 1, (1997), 230–258. doi:10.1214/aop/1024404287.
- [FOT11] M. FUKUSHIMA, Y. OSHIMA, and M. TAKEDA. *Dirichlet forms and symmetric Markov processes*, vol. 19 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, extended ed., 2011, x+489.
- [FS17] P. K. FRIZ and A. SHEKHAR. General rough integration, Lévy rough paths and a Lévy–Kintchine-type formula. *Ann. Probab.* **45**, no. 4, (2017), 2707–2765.
- [FV08] P. FRIZ and N. VICTOIR. On uniformly subelliptic operators and stochastic area. *Probab. Theory Related Fields* **142**, no. 3-4, (2008), 475–523. doi:10.1007/s00440-007-0113-y.
- [FV10] P. K. FRIZ and N. B. VICTOIR. *Multidimensional stochastic processes as rough paths*, vol. 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010, xiv+656.
- [GS77] V. GUILLEMIN and S. STERNBERG. *Geometric asymptotics*. American Mathematical Society, Providence, R.I., 1977, xviii+474 pp. (one plate). Mathematical Surveys, No. 14.
- [Lej06] A. LEJAY. Stochastic differential equations driven by processes generated by divergence form operators. I. A Wong–Zakai theorem. *ESAIM Probab. Stat.* **10**, (2006), 356–379. doi:10.1051/ps:2006015.
- [Lej08] A. LEJAY. Stochastic differential equations driven by processes generated by divergence form operators. II. Convergence results. *ESAIM Probab. Stat.* **12**, (2008), 387–411. doi:10.1051/ps:2007040.
- [LS99] T. LYONS and L. STOICA. The limits of stochastic integrals of differential forms. *Ann. Probab.* **27**, no. 1, (1999), 1–49. doi:10.1214/aop/1022677253.
- [Lyo98] T. J. LYONS. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14**, no. 2, (1998), 215–310. doi:10.4171/RMI/240.
- [Nas58] J. NASH. Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.* **80**, (1958), 931–954. doi:10.2307/2372841.
- [Osh13] Y. OSHIMA. *Semi-Dirichlet forms and Markov processes*, vol. 48 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2013, x+284. doi:10.1515/9783110302066.
- [SCS91] L. SALOFF-COSTE and D. W. STROOCK. Opérateurs uniformément sous-elliptiques sur les groupes de Lie. *J. Funct. Anal.* **98**, no. 1, (1991), 97–121. doi:10.1016/0022-1236(91)90092-J.
- [Str88] D. W. STROOCK. Diffusion semigroups corresponding to uniformly elliptic divergence form operators. In *Séminaire de Probabilités, XXII*, vol. 1321 of *Lecture Notes in Math.*, 316–347. Springer, Berlin, 1988. doi:10.1007/BFb0084145.
- [Stu95] K.-T. STURM. Analysis on local Dirichlet spaces. II. Upper Gaussian estimates for the fundamental solutions of parabolic equations. *Osaka J. Math.* **32**, no. 2, (1995), 275–312.

I. CHEVYREV, MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, ANDREW WILES BUILDING, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, UNITED KINGDOM

E-mail address: `chevyrev@maths.ox.ac.uk`

M. OGRODNIK, DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON, HUXLEY BUILDING, 180 QUEENS GATE, LONDON SW7 2AZ, UNITED KINGDOM

E-mail address: `marcel_ogrodnik@hotmail.com`